



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Serine Integrases

Citation for published version:

Merrick, CA, Zhao, J & Rosser, SJ 2018, 'Serine Integrases: Advancing Synthetic Biology', *ACS Synthetic Biology*. <https://doi.org/10.1021/acssynbio.7b00308>

Digital Object Identifier (DOI):

[10.1021/acssynbio.7b00308](https://doi.org/10.1021/acssynbio.7b00308)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

ACS Synthetic Biology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Serine Integrases: Advancing Synthetic Biology

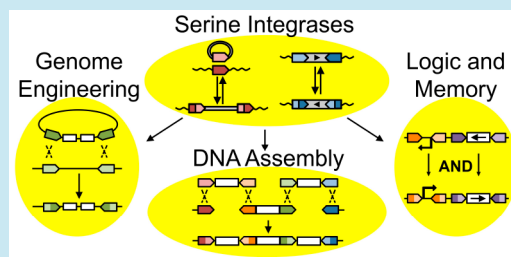
Christine A. Merrick,^{*,†} Jia Zhao,[‡] and Susan J. Rosser[†]

[†]School of Biological Sciences, University of Edinburgh, Roger Land Building, Alexander Crum Brown Road, Edinburgh EH9 3FF, U.K.

[‡]Novo Nordisk (China) Pharmaceuticals Co., Ltd., Lei Shing Hong Center, Guangshunnan Avenue, Beijing 100102, China

ABSTRACT: Serine integrases catalyze precise rearrangement of DNA through site-specific recombination of small sequences of DNA called attachment (*att*) sites. Unlike other site-specific recombinases, the recombination reaction driven by serine integrases is highly directional and can only be reversed in the presence of an accessory protein called a recombination directionality factor (RDF). The ability to control reaction directionality has led to the development of serine integrases as tools for controlled rearrangement and modification of DNA in synthetic biology, gene therapy, and biotechnology. This review discusses recent advances in serine integrase technologies focusing on their applications in genome engineering, DNA assembly, and logic and data storage devices.

KEYWORDS: serine integrase, recombination directionality factor (RDF), DNA assembly, genome engineering, logic devices, genetic data storage



DNA manipulation and modification are at the core of synthetic biology. Advances in the capacity to engineer DNA with increasingly complex properties are accelerating a diverse range of medical and biotechnological applications. In recent years, the principles of homologous recombination have been applied to DNA assembly¹ (for example, in construction of the first synthetic genome)² and to targeting specific changes in genomic DNA (for example, in nuclease-based genome editing with clustered regularly interspaced short palindromic repeat (CRISPR)-associated nucleases like Cas9 and Cpf1).^{3–8} However, another group of DNA modifying enzymes—the site-specific recombinases—are becoming established as vital tools for genetic modification both *in vivo* and *in vitro*. Unlike homologous recombination, site-specific recombination requires no extensive DNA sequence homology, no synthesis or degradation of DNA, and has no reliance on endogenous repair pathways or cofactors.

Serine integrases are a subfamily of serine site-specific recombinases, which are evolutionarily and mechanistically distinct from tyrosine site-specific recombinases.⁹ Serine recombinases make double strand breaks in DNA forming covalent 5'-phosphoserine bonds with the DNA backbone before religation, whereas tyrosine recombinases cleave single strands forming covalent 3'-phosphotyrosine bonds with the DNA backbone and rejoin the strands *via* a Holliday junction-like intermediate state. In comparison to tyrosine recombinases, recombination by serine integrases has the advantage of simplicity, normally only requiring the integrase protein and small *att* sites (<50 bp). This DNA recombination is also highly directional, reversed only in the presence of a single accessory protein called a recombination directionality factor (RDF), and requires no cofactors to bind or bend DNA. In comparison, the tyrosine integrase λ Int requires the host-encoded integration

host factor (IHF) protein to bend DNA for integration, and additional excisionase (Xis) and Fis proteins, encoded by the phage and host respectively, for DNA excision.¹⁰ Serine integrases can catalyze recombination between *att* sites on linear or circular DNA substrates and, depending on the position and orientation of the *att* sites, integrate, excise or invert sections of DNA (Figure 1).¹¹ In addition, serine integrases are active in a range of organisms including bacteria, yeast, flies, zebrafish, frogs, plants, and cell lines including mouse, rabbit, chicken, bovine, and human, reviewed elsewhere.^{12,13} The reader is referred to related reviews on the tyrosine and serine recombinase families,⁹ their development as tools,^{11,13} serine integrase structures and mechanisms,^{14,15} and their application in heterologous systems.^{12,13} This paper reviews established and emerging serine integrase technologies currently being applied to genome engineering, *in vitro* DNA assembly, and in the development of genetic logic and data storage devices.

SERINE INTEGRASE SITE-SPECIFIC RECOMBINATION

Serine integrases are encoded by temperate bacteriophages and catalyze their integration into bacterial genomes through recombination of *attP* (phage) and *attB* (bacteria) attachment sites, generating *attL* (left) and *attR* (right) sites. Upon lysogeny, the prophage is excised from the host chromosome by the serine integrase and its cognate RDF. Importantly, recombination of *attP* and *attB* is highly directional, it is not reversible in the presence of the serine integrase alone. The

Received: August 25, 2017

Published: January 9, 2018

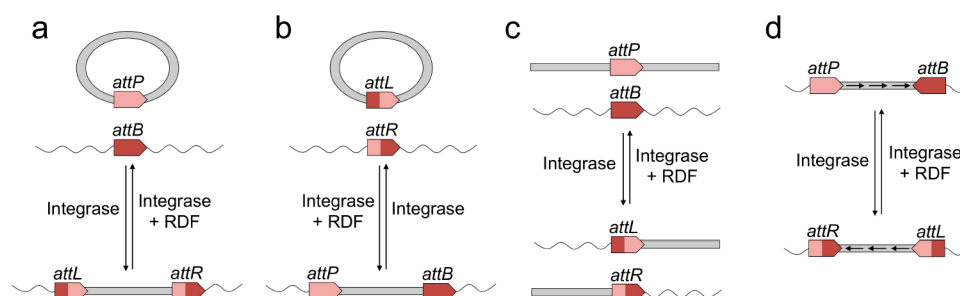


Figure 1. Examples of DNA rearrangement mediated by serine integrases and RDFs. In the presence of a serine integrase, *attP* and *attB* sites recombine to make *attL* and *attR* sites. In the presence of a cognate RDF, this reaction is reversed. For integration and excision of DNA, *att* sites are in the same orientation. For inversion of DNA, *att* sites are in opposite orientations. (a) DNA integration mediated by *attP* × *attB* recombination, and excision by *attL* × *attR* recombination. (b) Integration mediated by *attL* × *attR* recombination, and excision by *attP* × *attB* recombination. (c) Recombination between *att* sites in linear DNA. (d) DNA inversion through *attP* × *attB* recombination, and reversed by *attL* × *attR* recombination.

sequences of *attP* and *attB* are distinct from each other, and from *attL* and *attR* sites. Based on the crystal structure of the C-terminal domain of LI integrase bound to half of an *attP* site, Rutherford *et al.* proposed that the integrase adopts different spatial arrangements when bound to different *att* sites.¹⁶ From this, they formed a structure-based model for understanding why integrases preferentially form complexes with *attP* and *attB* sites, and prohibit *attL* × *attR* recombination.^{14–16}

To date, over 4000 putative serine integrases have been identified,¹⁷ 14 of which have been studied biochemically.¹⁴ Mechanistically, serine integrase dimers bind to *attP* and *attB* sites that are brought together to form a tetrameric synaptic complex (Figure 2). Serine residues cleave the *att* sites on

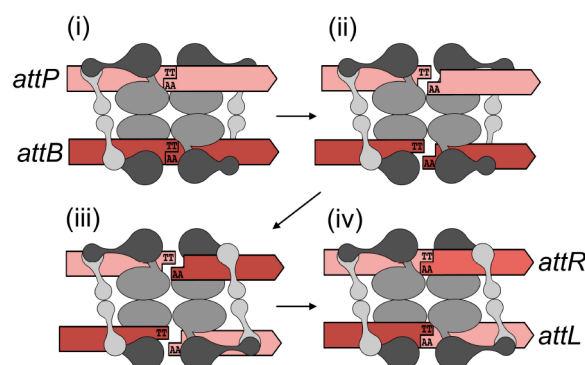


Figure 2. Mechanism of site-specific recombination by serine integrases. (i) A serine integrase binds its cognate *attP* and *attB* sites as a dimer before forming a protein tetramer in which the *att* sites are brought together. (ii) The four serine integrase protomers are activated to cleave the *att* sites generating half-sites with 3' overhangs, 2 base pairs in length. At this point each DNA overhang is attached to a protomer via a covalent 5' phosphoserine linkage. (iii) Two of the serine integrase subunits rotate in relation to the others swapping *attP* and *attB* half-sites. (iv) If the 3' overhangs of the half-sites are complementary, they will religate forming *attL* and *attR* sites. The C-terminal domains of serine integrase protomers are pale gray, *attP* half-sites are pale pink, and *attB* half-sites are dark pink.

either side of their central dinucleotide generating half sites with 3' 2 base pair overhangs. Next, integrase subunits rotate in relation to each other, effectively swapping half sites.^{18,19} Finally, if the 2 bp overhangs on the swapped half sites are complementary, they will religate to make *attL* and *attR* sites. Pairs of *attP* and *attB* sites with nonpalindromic central dinucleotide sequences that do not share inverted complementarity (TT, CT, GT, CA, CC, and TC) are considered to

be orthogonal. This means that in the presence of one serine integrase, as many as six pairs of these orthogonal *att* sites can recombine (*attP*^{TT} will specifically recombine with *attB*^{TT}, *attP*^{TC} will specifically recombine with *attB*^{TC}, and so on).^{20,21} Such orthogonality can also be applied to *attL* and *attR* sites.²⁰

RDFs are phage-encoded proteins that bind integrase proteins, stimulate recombination of *attL* and *attR* sites, and inhibit recombination of *attP* and *attB* sites.^{16,22–24} Each integrase subunit interacts with an RDF to trigger recombination of *attL* and *attR* sites, and it is hypothesized that this interaction induces conformational changes in the integrase that inhibits *attB* × *attP* recombination.^{16,25,26} Recently, these protein interactions and stoichiometries were optimized through fusion of phiC31 and Bxb1 integrases with their cognate RDFs, gp3 and gp47.²⁷ Among the nine RDFs identified to date, there is considerable variation in peptide length, sequence conservation, and location within their phage genome.^{22–24,28–33}

■ GENOME ENGINEERING WITH SERINE INTEGRASES

In the 30 years since serine integrases were first used to engineer streptomycete genomes at innate *attB* sites,^{34,35} these enzymes have been shown to be active in a number of organisms and have been engineered for enhanced *in vivo* activity.¹³ For example, enhanced phiC31 integrase activity was achieved in Chinese hamster ovary (CHO) cells with the addition of a C-terminal nuclear localization signal,³⁶ and a 3' polyadenylation signal has been adopted to stabilize phiC31 integrase mRNA injected into *Drosophila* embryos.³⁷ Chromosomal insertion of transgenes using serine integrases offers a number of advantages over alternative genome engineering methods. In comparison to homologous recombination-based methods including those that employ homing endonucleases such as zinc finger nucleases, transcription activator-like effector nucleases and CRISPR-Cas9, integration is mediated by only one enzyme without relying on host factors for DNA integration, and serine integrases do not degrade DNA or need to be engineered to bind a specific DNA sequence.³ Unlike transposons and retroviruses, integration can be targeted to a specific locus known to have minimal positional effects on transgene expression.³⁸ Transgenes integrated by serine integrases through *attP* × *attB* recombination cannot be inverted or remobilized without the presence of a cognate RDF. Genome engineering approaches that use serine integrases can be broadly grouped into those that integrate

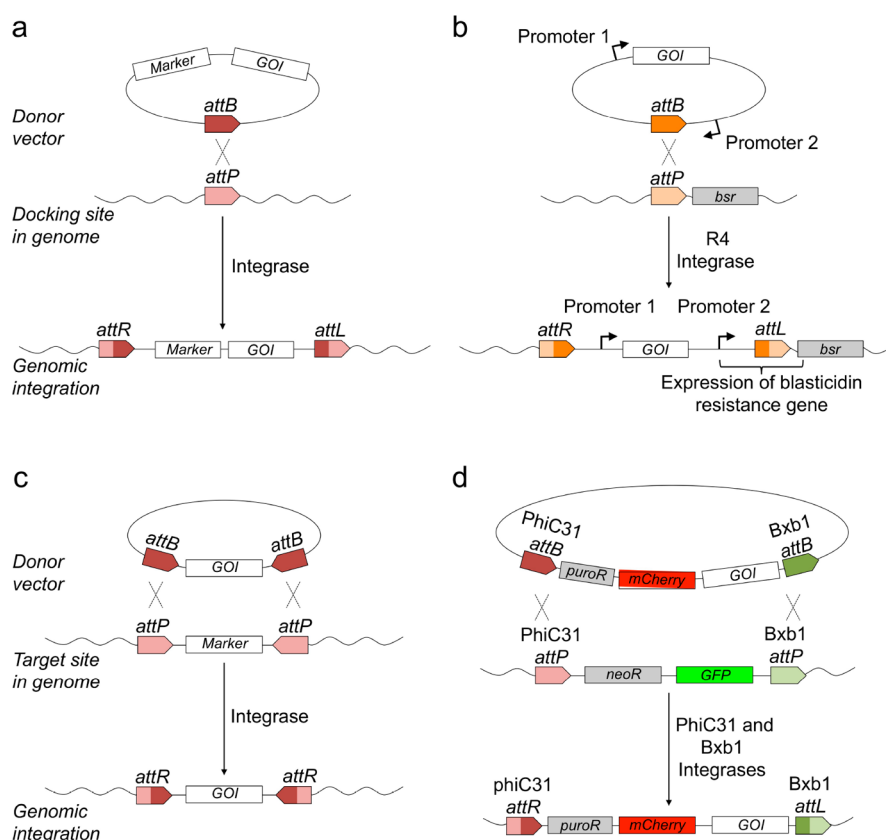


Figure 3. Serine integrase-mediated genome engineering strategies. (a) Genomic integration of DNA at a single *att* site with a serine integrase. A donor vector carries a gene of interest, a marker gene such as an antibiotic resistance gene, and an *attB* site. An *attP* site is present in the genome. The *attB* site in the donor vector recombines with an *attP* site in the genome in the presence of the integrase. This integrates the whole vector into this genomic locus where it is flanked by *attR* and *attL* sites. Integration of the vector can be detected by selection for the marker gene. (b) Jump-In GripTite Targeted Integration.⁵³ The R4 integrase *attP* site and a blasticidin resistance gene (*bsr*) are preintegrated in HEK293 genomic DNA at a known genomic locus. In this example, a donor vector carrying a gene of interest driven by Promoter 1, an R4 *attB* site, and Promoter 2, is integrated into the genome through *attP* × *attB* recombination in the presence of R4 integrase, expressed from a second vector. Once integrated, Promoter 2 drives expression of a blasticidin resistance gene enabling selection for integration events. (c) Recombinase-mediated cassette exchange (RMCE) with a serine integrase. A donor vector carries a gene of interest between two *attB* sites. A marker gene, such as one encoding a fluorescent protein, is present between two *attP* sites. Integrase mediates two *attP* × *attB* recombination reactions that result in exchange of the marker gene in the genome with the gene of interest. This exchange stops expression of the marker gene enabling screening for integration events. Once integrated, the gene of interest is flanked by two *attR* sites. (d) Dual integrase cassette exchange (DICE).⁶⁴ A DNA cassette carrying genes for neomycin resistance (*neo*) and green fluorescent protein (*GFP*) flanked by *attP* sites for ϕ C31 and Bxb1 integrases is inserted into the genome of embryonic stem cells or induced pluripotent stem cells by homologous recombination. A donor vector carries a puromycin resistance gene (*puroR*), a red fluorescent protein (*mCherry*), and a gene of interest flanked by ϕ C31 and Bxb1 *attB* sites. In the presence of ϕ C31 and Bxb1 integrases, expressed from separate expression vectors, the *attP* and *attB* sites recombine to remove the *neo*-*GFP* cassette from the genome, replacing it with the *puroR*-*mCherry*-*GOI* cassette flanked by ϕ C31 *attR* and Bxb1 *attL* sites. GOI stands for gene of interest, R4 *att* sites are orange, ϕ C31 *att* sites are pink, Bxb1 *att* sites are green, *attP* half-sites are pale, and *attB* half-sites are dark.

DNA into pre-existing genomic loci (pseudosites), those that use a “landing pad” (a single *att* site integrated into the genome), and recombinase-mediated cassette exchange (RMCE).

Genomic sequences that bear some similarity to wild-type *att* sites, termed pseudo *att* sites, facilitate integration of foreign DNA. Gene integration into pseudo *att* sites in the presence of ϕ C31 integrase has been demonstrated in human, mouse and rat genomes for the development of gene therapies^{39–43} and to insert DNA for reprogramming of adult mouse fibroblasts into induced pluripotent stem cells.^{44,45} The existence of pseudo *attP* sites has been the subject of some debate due to low levels of sequence homology between native and pseudo *attP* sites, and observations of background integration into nonspecific loci.^{40,46–48} However, a bioinformatics approach has identified a ~30 bp consensus motif containing inverted repeats found to

be present in the native ϕ C31*attP* sequence, in the 19 pseudo *attP* sites with the highest integration frequencies in the human genome, and in most of the pseudo *attP* sites with lower integration frequencies, providing indirect evidence for pseudo *attP* sites.⁴⁹ The main shortcoming associated with pseudo sites is that DNA integration cannot be specifically targeted to one individual pseudo *att* site making it difficult to make directly comparable strains.

A landing pad is an *att* site that has been integrated into a genome by transposition or homologous recombination. Genetic cargo that is to be inserted into the landing pad is provided on a vector that contains a complementary *att* site. In the presence of a cognate serine integrase the *att* sites recombine, inserting the plasmid into the genome where it is then flanked by *att* sites⁴⁶ (Figure 3a). In the *Drosophila* research community, collections of strains with landing pads at

different genomic loci have been made and characterized for positional effects on transgene expression,^{37,50,51} and integration of a large vector (133 kb) has been achieved in the presence of phiC31 integrase expressed from injected mRNA.⁵⁰ For mammalian cells, the Jump-In Targeted Integration System from ThermoFisher Scientific offers CHO-K1 and human embryonic kidney (HEK) 293 cells with preintegrated *attP* sites for R4 integrase at known genomic loci.⁵² A donor vector, carrying a complementary *attB* site and a gene of interest, is integrated into the genome in the presence of R4 integrase, expressed from a second vector. GripTite is a version of Jump-In in which the expression vector carries a promoter to activate an antibiotic resistance gene located in the target genome next to the *attP* site (Figure 3b).⁵³ A benefit of landing pads is that the genomic locus is known prior to insertion of transgenes. However, as is the case for pseudo sites, a strategy is required to detect integration, often requiring incorporation of extra DNA for selection, and an additional strategy may also be required to excise excess donor vector DNA from the genome.

RMCE is a targeted genome integration strategy originally established with tyrosine recombinases in which two recombination events mediate integration of a desired transgene and excision of the donor vector backbone (Figure 3c).^{54,55} In RMCE, a target site consisting of a selection marker flanked by two noncomplementary *att* sites is inserted at a genomic locus through transposition or homologous recombination. A vector containing the gene of interest flanked with *att* sites complementary to those in the target site undergoes two recombination reactions with the target site in the presence of the recombinase. Serine integrases are ideal candidates to mediate RMCE because recombination of *attP* and *attB* sites is highly directional, the *att* sites are small, and transgenes can be integrated from linearized or supercoiled DNA. In comparison experiments, Bxb1 and phiBT1 integrases were the most efficient serine integrases for RMCE in human HT1080 cells and *Saccharomyces cerevisiae*, respectively.^{56,57} The *Drosophila* research community has used a phiC31 integrase-mediated RMCE approach extensively to study transgene expression from defined genomic loci.^{58–63}

Dual integrase cassette exchange (DICE), developed by the Calos group, is a variation of RMCE that uses a pair of orthogonal serine integrases⁶⁴ (Figure 3d). The group used a bioinformatics approach to identify a suitable locus for transgene expression, and then homologous recombination to insert a target sequence containing genes for antibiotic resistance and GFP flanked by *attP* sites for phiC31 and Bxb1 integrases. Using these enzymes they integrated transgenic DNA flanked with complementary *attB* sites, and were able to rapidly generate libraries of embryonic stem cells and induced pluripotent stem cells carrying different combinations of neuronal transcription factor genes. The DICE reactions were invariably precise and transgene insertion was always in the same orientation.⁶⁴ In addition, the same group demonstrated that the phiC31 integrase RDF could mediate chromosomal excision through recombination of *attL* and *attR* sites in HeLa-derived cell lines.⁶⁵

These studies are of particular significance because combinations of serine integrases and RDFs could, in future, be used for serial RMCEs at the same genomic locus. Also, in the advent of genome synthesis projects, combinations of serine integrases could be applied to the rearrangement and rapid evolution of synthetic genomes in an approach inspired by Synthetic Chromosome Rearrangement and Modification by

LoxPsym-mediated Evolution (SCRaMbLE).^{66,67} SCRaMbLE, which was incorporated into the synthetic yeast genome project, Sc2.0, uses an engineered Cre recombinase and nondirectional *loxP* sites dispersed throughout synthetic chromosomes to invert and delete sections of DNA allowing the formation of structurally distinct genomes.^{68,69} Not only will this rapid evolution approach contribute to understanding of genome structures, but strains could also be isolated with enhanced robustness for industrial processes such as fermentation. A similar system, developed with serine integrases, could utilize RDFs for targeted postrearrangement modification of specific loci.

■ DNA ASSEMBLY BY SITE-SPECIFIC RECOMBINATION

Since the introduction of BioBricks 15 years ago,⁷⁰ innovative DNA assembly strategies have been fundamental to progress in synthetic biology including construction of the genome of the first synthetic organism,² metabolic engineering of microbes for production of high-value compounds,⁷¹ establishment of DNA registries for the distribution of reusable DNA parts,⁷² and cloning of repetitive DNA sequences.^{73,74} Current DNA assembly methods can be grouped into three categories: endonuclease-mediated assembly methods including BioBrick-based methods,^{70,75,76} and Golden Gate-based Assembly;^{77–79} homology-based methods including sequence and ligation independent cloning (SLIC),⁸⁰ Gibson Assembly,⁸¹ DNA Assembler in yeast,⁸² and PCR-based methods like circular polymerase extension cloning (CPEC)⁸³ and PaperClip;⁸⁴ and site-specific recombination-based methods including Gateway,^{85–87} site-specific recombination-based tandem assembly (SSTRA)⁸⁸ and serine integrase recombinational assembly (SIRA).²⁰ The main complication of endonuclease-based methods is the requirement to remove incompatible restriction sites from DNA parts prior to assembly, meaning that often DNA parts must be synthesized, or precloned and mutated. Limitations of *in vitro* homology-based methods are associated with the size of DNA parts: both SLIC and Gibson Assembly use exonuclease activity to generate single-strand complementary overhangs on DNA parts meaning they have the potential to degrade small DNA parts; and PCR-based methods are limited to DNA parts that are short enough to be amplified *in vitro* with a DNA polymerase. Although these methods form no defined scar sequences, mutation events can occur at junctions between DNA parts, particularly in repetitive sequences. Also, once assembled, the junctions between DNA parts do not allow individual sections of DNA to be easily exchanged.

Site-specific recombination-based DNA assembly methods recombine *att* sites that flank DNA parts. Gateway Cloning uses the tyrosine integrase from phage λ , together with an accessory protein, IHF, to clone DNA parts flanked with orthogonal *attB* sites into donor vectors containing complementary *attP* sites, through *attP* \times *attB* recombination.⁸⁵ This produces entry vectors in which DNA parts are flanked with *attL* sites. From this point, the DNA part can be transferred into expression vectors containing complementary *attR* sites through recombination mediated by λ integrase, IHF and an excisionase protein that reverses the direction of the recombination reaction, enabling *attL* \times *attR* recombination. Additional orthogonal *att* sites have enabled Gateway assembly of up to five DNA parts in a single reaction and, due to the controllable directionality of the recombination reaction, it is possible to modify a Gateway construct postassembly.⁸⁷ However, pre-cloning of DNA parts

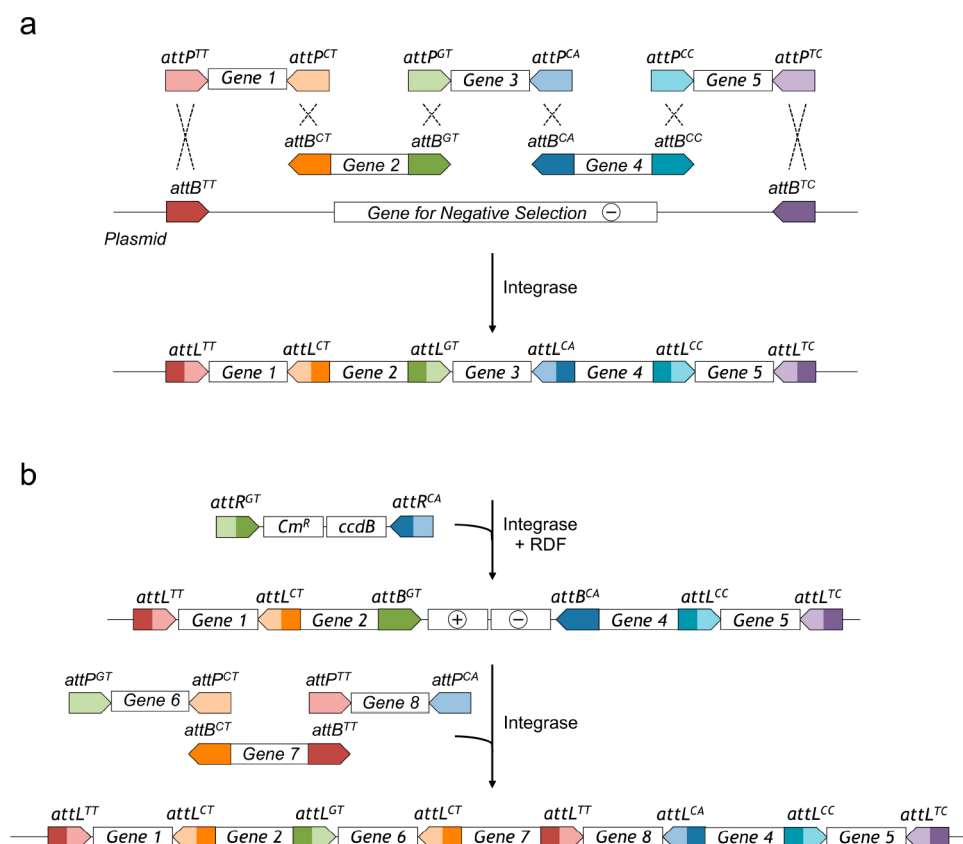


Figure 4. Serine integrase recombinational assembly (SIRA). (a) SIRA for assembling five DNA parts in a specific, predefined order in a plasmid by *attP* × *attB* recombination. Five linear DNA parts containing *genes 1* to *5* are flanked with *attB* or *attP* sites in alternating orientations using PCR. The *attP* and *attB* sites have orthogonal central dinucleotide sequences so that an *attB^{TT}* site will recombine with an *attP^{TT}* site, an *attP^{CT}* site will recombine with a *attB^{CT}* site, and so on. This ensures the DNA parts assemble in a specific order. The SIRA substrate plasmid contains a gene for negative selection between two inverted orthogonal *att* sites. Here, the *ccdB* gene is between *attB^{TT}* and *attB^{TC}*. The *ccdB* gene encodes a cytoprotein that is toxic in *E. coli*. In the assembly reaction, the orthogonal *attP* and *attB* sites with matching central dinucleotides recombine replacing the *ccdB* gene with DNA parts in the gene order 1–2–3–4–5. This enables selection for assembled products upon transformation of cells because substrate plasmids harboring the *ccdB* gene will cause cell death.²⁰ (b) Targeted postassembly modification of a SIRA construct through *attL* × *attR* recombination. A linear piece of DNA carrying genes for positive and negative selection (here, chloramphenicol resistance (*Cm^R*) and *ccdB*, respectively) between *attR^{GT}* and *attR^{CA}* recombines with *attL^{GT}* and *attL^{CA}* in the SIRA construct containing *genes 1* to *5* in the presence of integrase and its cognate RDF. In this example, this piece of DNA known as a “secondary insertion site” replaces *gene 3* and, as a result, is flanked by *attB^{GT}* and *attB^{CA}*. The chloramphenicol resistance gene facilitates selection for plasmids containing the secondary insertion site. Further DNA parts, here carrying *genes 6*, *7*, and *8*, can be assembled in the place of the secondary insertion site through *attP* × *attB* recombination. This removal of the secondary insertion site including the *ccdB* gene enables selection for assembled products upon transformation of cells.²⁰ In (a) and (b) *attP*, *attB*, *attL* and *attR* sites with complementary central dinucleotides have matching colors, *attP* half-sites are pale and *attB* half-sites are dark. Symbols “+” and “–” indicate positive and negative selection genes.

into donor and entry vectors is time-consuming, and the *att* sites to which IHF must bind and bend are large—the Gateway *attP* site is 200 bp.⁸⁹

Compared to those recognized by tyrosine integrases, *att* sites for serine integrases are small (<50 bp), and require no accessory proteins to bind or bend them for recombination. It is therefore possible to recombine serine integrase *att* sites on linear pieces of DNA, eliminating the need to preclone DNA parts into entry vectors. Zhang *et al.* developed SSTRA which uses linearized DNA parts flanked with an *attB* site and an *attP* site oriented in the same direction and mutated to recombine orthogonally.⁸⁸ SSTRA has been used to assemble a functional carotenoid biosynthetic pathway, and a 56 kb construct of DNA parts ~10 kb in size, each interspersed with orthogonal *attL* sites in the same orientation.^{88,90}

In an independent study, Colloms *et al.* developed SIRA, in which linear DNA parts are flanked with orthogonal *attP* and *attB* sites in alternating orientations with different central

dinucleotides.²⁰ SIRA incorporates a set of simple design principles to enhance DNA assembly by prohibiting intramolecular recombination (Figure 4a).^{20,91} Using phiC31 integrase, it was demonstrated that multipart assembly with SIRA could be applied to make combinatorial libraries of constructs for rapid optimization of metabolic pathways by assembling parts in multiple gene orders in a one-pot reaction, and by employing degenerate ribosome binding sites. SIRA can be used to optimize gene expression levels and identify bottlenecks in metabolic pathways. Once assembled, DNA parts in a SIRA construct are interspersed with *attL* sites in alternating orientations and with different central dinucleotides. This facilitates targeted postassembly modification of an assembled construct for replacement and addition of more DNA parts (Figure 4b). A key feature of SIRA is that its design principles can be applied to other serine integrases that have the potential to be used in combination for DNA assembly and modification *in vitro* or *in vivo*.⁹²

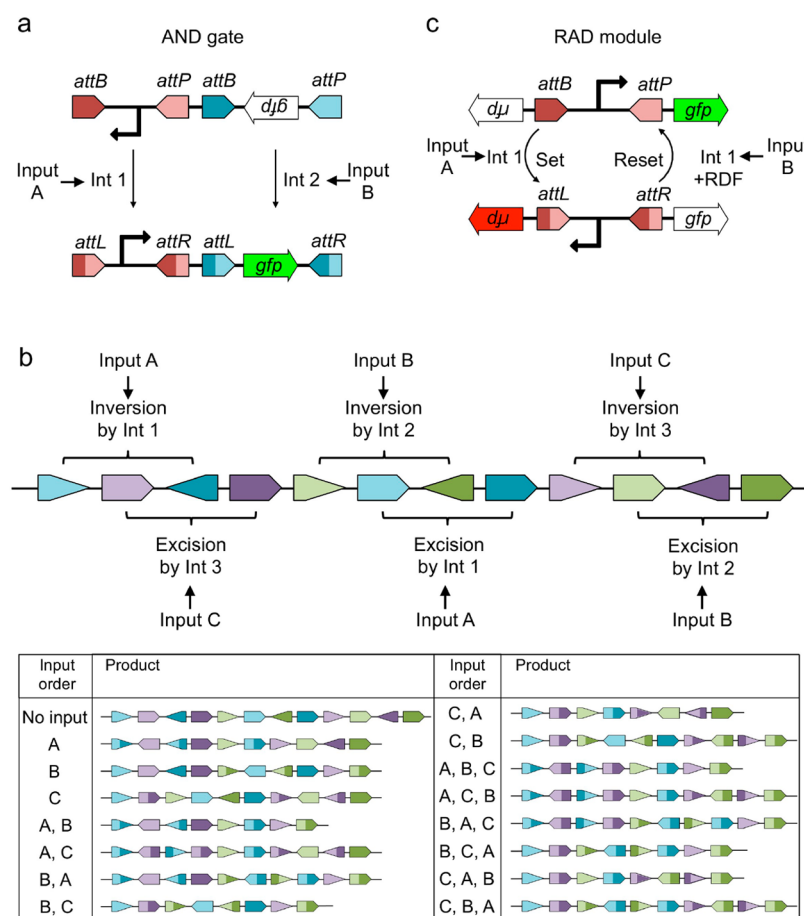


Figure 5. Serine integrase logic and memory devices. (a) The AND gate based on flipping a promoter and a gene. Int 1 (phiC31 integrase) and Int 2 (Bxb1 integrase) are expressed from promoters that are activated by input signals A and B, respectively. After both input signals have been present, the expressed integrases recognize and recombine their cognate sites and flip the promoter and the *gfp* gene to the same orientation, resulting in the expression of the reporter protein, GFP.⁹³ (b) The framework of recombinase-based state machines (RSMs). A RSM requires a DNA register consisting of orthogonal and interleaving pairs of *att* sites, and their corresponding integrases which are expressed in response to input signals. Three chemical signals are used to independently trigger expression of integrases: Input A for Int 1 (Bxb1 integrase); Input B for Int 2 (TP901 integrase); Input 3 for Int 3 (A118 integrase). The integrases rearrange the DNA register by inverting and excising sections of DNA. The table shows that each possible temporal order of input signals results in a DNA register with a unique sequence.²¹ (c) The rewritable recombinase addressable data (RAD) module. Int 1 (Bxb1 integrase) recombines *attP* and *attB* sites in opposite orientations to generate *attL* and *attR* sites. This inverts the intervening constitutive promoter, switching on expression of RFP and switching off expression of GFP. In the presence of the RDF (gp47), Bxb1 integrase recombines the *attL* and *attR* sites to generate *attB* and *attP* sites. This restores the constitutive promoter to the original state, switching on expression of GFP and switching off expression of RFP.⁹⁶ The *att* sites are blue for Bxb1 integrase; pink for phiC31 integrase; green for TP901 integrase; and purple for A118 integrase. Darker shades indicate *attB* half sites, and lighter shades indicate *attP* half sites. Different shapes are used to distinguish between *att* sites with orthogonal dinucleotides.

While it has been demonstrated that SSTR and SIRA are reliable methods for cloning genes and assembling metabolic pathways,^{20,88,90} both create constructs interspersed with *att* sites deeming them unsuitable for applications such as directed evolution of proteins. It is also worth noting that low levels of recombination have been observed between orthogonal *att* sites under selective pressure,²⁰ and that it is possible that *att* site within constructs could affect gene expression levels.^{93,94}

SERINE INTEGRASE LOGIC AND MEMORY DEVICES

Much work in synthetic biology has been inspired by the goal to create biological systems that can record specific input signals, perform computations, and respond with outputs that can be applied in biotechnological, environmental and therapeutic contexts. To this end, a number of biological components including logic and memory devices have been

developed based on transcriptional modules, RNA-based post-transcriptional regulation, and protein interactions (reviewed elsewhere^{21,95}). However, in order to maintain a steady state, these systems rely on levels of transcription that could impose considerable resource burden on the host cell by redirecting nucleic acids, amino acids and other metabolites away from native cellular processes. This could cause evolutionary counter-selection of such systems, or render them subject to changes in cellular metabolism resulting in spontaneous alterations in device outputs.⁹⁶ As applications of these biological computation systems become more complex, so too does the requirement for robust, scalable components that are genetically stable and cause minimal resource burden.

Recombinase-based logic and memory devices respond to inputs by rearranging DNA. This means that the presence of an input can be recorded as a permanent, heritable DNA rearrangement from a single pulse of recombinase expression,

and therefore causes minimal resource burden. Such systems have been described using tyrosine recombinases such as FimE invertase which irreversibly flips the orientation of specified DNA sequences,⁹⁷ Cre and Flp recombinases,^{98,99} and combinations of tyrosine and serine recombinases.¹⁰⁰ A system recently developed in mammalian cells, Boolean logic and arithmetic through DNA excision (BLADE),¹⁰⁰ uses tyrosine recombinases to excise transcription terminators located upstream of genes, and serine recombinases to invert genes downstream of a stationary promoter, analogous to BUF (buffer) gates. Using an array of four of these BUF gates, Weinberg *et al.* built a field-programmable, read-only memory circuit that can be programmed with combinations of four recombinases to act as any of 16 2-input Boolean logic gates.¹⁰⁰

In earlier work, Siuti *et al.* elegantly demonstrated the potential of using simply ϕ C31 and Bxb1 serine integrases to carry out logic functions in *E. coli* by constructing all 16 two-input Boolean logic gates in DNA with promoters, terminators and genes that could be inverted by just *attP* \times *attB* recombination⁹³ (Figure 5a). The AND gate maintained a stable output memory for over 90 generations after input signals had been removed, and this output could be detected by PCR after cell death. Furthermore, this study produced a suite of two-bit digital-to-analog converters in which two *gfp* expression cassettes could be flipped by two integrases, respectively, to be expressed from different constitutive promoters of different strengths. Different combinations of two inducers (digital input) triggered expression of either, both or no integrases, and therefore produced four different levels of GFP expression ranging from low to high (analog output).⁹³ Independently, the Endy group developed permanent, amplifying AND, NAND, OR, NOR, XOR, and XNOR transcriptor-based gates using two serine integrases.⁹⁴ These used the inversion or excision functions of TP901 and Bxb1 integrases to rearrange the orientations of regulatory DNA sequences, thereby controlling the flow of RNA polymerase along DNA and thus amplifying expression of a reporter gene.⁹⁴ In recognition of the fact that a small change in input signal can trigger an all-or-nothing effect on recombination and therefore output, the group endeavored to compare gate output levels (GFP expression) to increasing levels of input signals (inducers of integrase expression) to find a combination of input levels that would provide a threshold above which GFP expression was “on” and below which it was “off”. By producing binary outputs (*i.e.*, expression of GFP is either switched on or off) in response to low or high inputs, these transcriptor gates act as analog-to-digital converters.⁹⁴ One such transcriptor-based AND gate was later shown to enable a bacterial whole-cell biosensor to execute a signal-processing algorithm in clinical samples demonstrating a new approach to medical diagnosis.¹⁰¹

In recent years, a number of studies have contributed to increasing the information storage capacity of integrase-based logic and memory devices. Using a bioinformatics approach, Yang *et al.* mined 4105 putative large serine-type phage integrases from prophage genomes and then, using *att* sites for 11 integrases that shared a maximum of 60% sequence identity, built an 11-bit array consisting of 11 distinct segments of DNA that could be flipped in orientation.¹⁷ This DNA array was 2 kb in size and capable of storing 1.375 bytes of data. In addition to recording the presence of biological inputs, memory devices with overlapping switch modules for different integrases can record the temporal order of these input signals. This was first realized when the inversion systems of *fim* and *hin* tyrosine and

serine recombinases were integrated to create a heritable sequential memory switch with as many as four states.¹⁰² More recently, the Lu group created recombinase-based state machines (RSMs) which use the inversion and excision functions of multiple input-driven serine integrases to rearrange DNA registers composed of orthogonal and interleaving pairs of *att* sites (Figure 5b).²¹ In this framework, DNA registers are rearranged into completely different conformations depending on the temporal order of input signals. Using TP901, Bxb1, and A118 integrases, and multiple orthogonal pairs of *att* sites, Roquet *et al.* demonstrated that the information capacity of this framework could be scaled to that of a 16-state RSM and, by incorporating regulatory elements and fluorescent protein genes, built functional 16-state gene-regulatory RSMs (GRSMs).²¹

An intrinsic issue of logic and memory devices that rely on the switching of states is noise generated by heterogeneity in biological populations. Taking this into account, Hsiao *et al.* developed a two-input, serine integrase-based temporal logic gate containing *att* sites for Bxb1 and TP901–1 integrases, which was integrated into the chromosome of a strain of *E. coli*.¹⁰³ By analyzing the heterogeneity of individual cellular states generated in response to different sequences of chemical inputs, applying a stochastic model of single-cell trajectories, and amalgamating these into population-level distributions, this group was able to determine the order, timing and duration of input signals.¹⁰³ A mechanistic model of this two-input temporal logic gate has been developed.¹⁰⁴

The serine-integrase logic and memory devices discussed up to this point rearrange DNA to produce stable, heritable, permanent outputs. While such devices are ideal for recording one-off historical events such as the presence of an environmental signal, rewritable devices are being developed to enable counting of real-time, dynamic biological processes such as cell divisions. The ability to stringently control the direction of recombination by serine integrases with RDFs makes them ideal mediators to set and reset the orientation of DNA in a controlled manner, thus increasing the data storage capacity of memory devices and enabling reuse of logic gates. The Endy group applied serine integrases and their cognate RDFs to data storage by making a rewritable recombinase addressable data (RAD) module⁹⁶ (Figure 5c). In this module, the orientation of a DNA sequence flanked by *attP* and *attB* sites was flipped by Bxb1 integrase setting the module to “state 1” in which the DNA sequence is flanked by *attL* and *attR* sites. Upon induction of the cognate RDF, gp47, the DNA sequence was reset to its original orientation, state “0”, by *attL* \times *attR* recombination.⁹⁶ A mechanistic model of a rewritable RAD module that uses ϕ C31 integrase and its cognate RDF, gp3, has been described.¹⁰⁵ It has been proposed that an N-bit counter built from 2N RAD modules could count up to 2^N states, limited only by the number of known orthogonal serine integrases with cognate RDFs.¹⁰⁶

In contrast to recombinase-based memory devices, a number of recent studies have adopted CRISPR-Cas9 systems to record the presence of biological signals. Temporal recording in arrays by CRISPR expansion (TRACE) uses CRISPR as a biological tape to record cellular events over time.¹⁰⁷ In this system, detection of a biological signal triggers an increase in the copy number of a reporter plasmid. The increase in the presence of this plasmid is recorded as increased incorporation of corresponding spacers into CRISPR arrays across a population of cells. However, data storage capacity is limited by the slow

rate of spacer incorporation (~ 1 in 15 arrays incorporated a new spacer each day) and an analytical model is required to interpret CRISPR sequences across the population of cells.¹⁰⁷ Mammalian synthetic cellular recorders integrating biological events (mSCRIBE) from the Lu lab,¹⁰⁸ and homing CRISPR barcodes from the Church lab,¹⁰⁹ use self-targeting guide RNAs (stgRNAs)¹⁰⁸ (or homing guide RNAs¹⁰⁹), which direct a Cas9 endonuclease to cleave loci that encode the stgRNAs. This induces nonhomologous end joining (NHEJ), which introduces mutations to the stgRNAs. Over time, these accumulating mutations form a record of analog data, such as fluctuations in biological signals, and computational metrics are applied to determine the duration and magnitude of the input signal, or to trace cell lineage. Limitations of these approaches are associated with *in situ* sequencing for analysis of a mixed population of cells and the restricted diversity that can be generated in the guide RNAs. Some mutations in these systems occur with a higher frequency than others,¹⁰⁸ and NHEJ can repair DNA cleavage without introducing mutations at all. It is also not known if Cas9 would have any off-target activity, or if integrated arrays of nuclease targets would affect genome stability.

Compared to transcription-based biological logic and memory systems including those that use CRISPR-Cas9 components, serine integrase-based devices have a simple composition comprising of DNA, integrases and RDFs. As serine integrases and RDFs are active in a range of cell-types and do not rely on endogenous DNA repair systems for DNA modification, these devices could be made portable and utilized in multiple organisms. Data stored in DNA can be reliably retained throughout cell divisions, transferred across species, and easily detected after cell death. Importantly, these devices have the potential to detect transient signals and to record data in otherwise inaccessible environments.

■ PROSPECTS FOR SERINE INTEGRASES

The capacity of serine integrases to recombine DNA both *in vitro* and in a broad spectrum of *in vivo* systems, and to record the presence of biological signals and carry out logic functions, means they have the future potential to be applied in self-programming, decision-making biological machines. These machines could be applied in disease diagnosis, production and delivery of bespoke therapeutics, remote environment monitoring, universal bioremediation systems, and self-evolving microbial factories for biosynthesis of biomaterials and valuable compounds.

Taking into account some of the serine integrase-based systems discussed in this review, one could consider how such a machine could take shape. The RSM described above uses a DNA register composed of two pairs of overlapping and orthogonal *att* sites for three different integrases, that produces 16 different DNA sequence outputs in response to 16 different temporal input sequences²¹ (Figure 5b). If four such DNA registers, each responsive to three different integrases, were implemented side by side, this would generate a 12-integrase RSM with 65 536 potential temporal input sequences and the same number of DNA sequence outputs. Given that 11 serine integrases have been identified that recombine their own *att* sites without showing cross-reactivity with *att* sites of other integrases,¹⁷ it is not unreasonable to imagine this and larger RSMs will be constructed in the near future. In the long term, synthetic genomes could be constructed in a RSM-based structure to enable rapid evolution in a random manner similar

to SCRaMBLE^{66,67} (all integrases triggered at once), or in a targeted, finely tuned manner (in response to a distinct temporal sequence of inputs). It is also realistic to expect that RDFs will become key components of RSM-like systems for rewritable logic and counting functions, and for incorporation of additional DNA sequences. This would utilize processes not dissimilar to targeted postassembly modification with SIRA (targeting pairs of *att* sites with different central dinucleotides) (Figure 4b), DICE (targeting two *att* sites for different integrases) (Figure 3d), and DNA integration into a single *att* site (Figure 3a).

Progress toward realization of such multifunctional systems is largely restrained by one key bottleneck: the current collection of serine integrase-related tools is small. To the best of our knowledge, there are currently 14 biochemically characterized serine integrases,¹⁴ nine RDFs have been identified,^{22–24,28–33} and each integrase has a potential set of six orthogonal *att* sites.²⁰ The first steps toward expanding this collection of tools would be to characterize the 4105 recently identified putative large serine-type phage integrases, and to develop bioinformatic solutions to facilitate mining of phage genomes for potential RDFs (this is currently a challenge because known RDFs differ considerably in sequence, size and genomic location). Ideally, each integrase and RDF characterized should be assessed for cross-reactivity with *att* sites corresponding to other integrases and RDFs. Logical progression from integrase and RDF discovery would be generation of more integrase-RDF chimeras to increase the stringency of *attL* \times *attR* recombination reactions.²⁷

There are a number of reasons in favor of expanding the repertoire of available *att* sites. First, the more orthogonal *att* sites an integrase could target, the more scalable its applications would be. Second, it has been noted that some *att* sites could have an effect on mRNA stability or transcription initiation rates of reporter genes.^{93,94} Finally, as previously mentioned, selective pressure can result in low levels of cross-reactivity between orthogonal *att* sites with different central dinucleotide sequences.²⁰ Generation of additional orthogonal *att* sites would likely entail engineering of serine integrases to recognize and bind to new DNA sequences. Farruggio *et al.* created phiBT1-phiC31 and phiC31-TG1 hybrid integrases by joining their N-terminal and C-terminal portions, and these were active in *E. coli* on *att* sites that were hybrids between those of the donor enzymes.¹¹⁰ It is currently not known if recombination of these *att* sites could be reversed in the presence of an RDF. Several “hyperactive” small serine recombinase mutants that act independently of their native recombination regulatory systems have been engineered to recognize non-native DNA sequences through replacement of their C-terminal DNA-binding domains with DNA-binding zinc finger domains, or complexes of dead Cas9 and guide RNAs.^{111–114} In contrast, the C-terminal domains of serine integrases are much larger than those of small serine recombinases, and consist of multiple structural domains that are involved in regulating *att* site recognition and recombination directionality.^{16,115,116} Given that deletion of the coiled-coil motif from the C-terminal of LI integrase renders DNA integration by *attP* \times *attB* recombination negligible, and facilitates excision through *attP* \times *attB*, *attL* \times *attR*, *attP* \times *attP*, and *attB* \times *attB* recombination,¹⁶ it is plausible that replacing serine integrase C-terminal domains with alternative DNA binding strategies would compromise their functionality.

In conclusion, serine integrases, and their cognate RDFs and *att* sites, provide reliable DNA recombination systems that can be used in a number of applications including *in vivo* genomic integration of DNA in a range of organisms, and *in vitro* assembly and targeted modification of DNA constructs. These recombination systems have also been applied in low-burden, heritable DNA logic and memory devices that can store information permanently or in a rewritable fashion. As many serine integrases work independently on distinct *att* sites, there is potential to scale up their applications by using multiple recombination systems together. This is currently limited to the relatively small number of biochemically characterized serine integrases, RDFs and *att* sites available. Future advances in serine integrase technologies would benefit greatly from development of new strategies for identification and characterization of these components, and greater understanding of how they interact and regulate recombination.

AUTHOR INFORMATION

Corresponding Author

*E-mail: christine.merrick@ed.ac.uk. Tel: 0131 650 5416.

ORCID

Christine A. Merrick: 0000-0002-6464-1224

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Funding was provided by the UK Engineering and Physical Sciences Research Council (EP/H019154/1 and EP/K034359/1).

ABBREVIATIONS

RDF, recombination directionality factor; CRISPR, clustered regularly interspaced short palindromic repeat; RMCE, recombinase-mediated cassette exchange; CHO, Chinese hamster ovary; HEK, human embryonic kidney; DICE, Dual integrase cassette exchange; SCRaMBLE, Synthetic Chromosome Rearrangement and Modification by LoxPsym-mediated Evolution; SLIC, sequence and ligation independent cloning; CPEC, circular polymerase extension cloning; SSTR, site-specific recombination-based tandem assembly; SIRA, serine integrase recombinational assembly; BLADE, Boolean logic and arithmetic through DNA excision; RAD, recombinase addressable data; RSM, recombinase-based state machines; GRSM, gene-regulatory recombinase-based state machines; TRACE, temporal recording in arrays by CRISPR expansion; mSCRIBE, mammalian synthetic cellular recorders integrating biological events; stgRNA, self-targeting guide RNAs; NHEJ, non-homologous end joining

REFERENCES

(1) Casini, A., Storch, M., Baldwin, G. S., and Ellis, T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.* 16, 568–576.

(2) Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., Smith, H. O., and Venter, J. C. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (Washington, DC, U. S.)* 329, 52–56.

(3) Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405.

(4) Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771.

(5) Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (Washington, DC, U. S.)* 337, 816–821.

(6) Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science (Washington, DC, U. S.)* 339, 819–823.

(7) Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-guided human genome engineering via Cas9. *Science (Washington, DC, U. S.)* 339, 823–826.

(8) Murovec, J., Pirc, Ž., and Yang, B. (2017) New variants of CRISPR RNA-guided genome editing enzymes. *Plant Biotechnol. J.* 15, 917–926.

(9) Grindley, N. D., Whiteson, K. L., and Rice, P. A. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* 75, 567–605.

(10) Landy, A. (1989) Dynamic, structural, and regulatory aspects of λ site-specific recombination. *Annu. Rev. Biochem.* 58, 913.

(11) Olorunniji, F. J., Rosser, S. J., and Stark, W. M. (2016) Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem. J.* 473, 673–684.

(12) Baltz, R. H. (2012) Streptomyces temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms). *J. Ind. Microbiol. Biotechnol.* 39, 661–672.

(13) Fogg, P. C., Colloms, S., Rosser, S., Stark, M., and Smith, M. C. (2014) New applications for phage integrases. *J. Mol. Biol.* 426, 2703–2716.

(14) Van Duyne, G. D., and Rutherford, K. (2013) Large serine recombinase domain structure and attachment site binding. *Crit. Rev. Biochem. Mol. Biol.* 48, 476–491.

(15) Rutherford, K., and Van Duyne, G. D. (2014) The ins and outs of serine integrase site-specific recombination. *Curr. Opin. Struct. Biol.* 24, 125–131.

(16) Rutherford, K., Yuan, P., Perry, K., Sharp, R., and Van Duyne, G. D. (2013) Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res.* 41, 8341–8356.

(17) Yang, L., Nielsen, A. A., Fernandez-Rodriguez, J., McClune, C. J., Laub, M. T., Lu, T. K., and Voigt, C. A. (2014) Permanent genetic memory with > 1-byte capacity. *Nat. Methods* 11, 1261–1266.

(18) Bai, H., Sun, M., Ghosh, P., Hatfull, G. F., Grindley, N. D., and Marko, J. F. (2011) Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7419–7424.

(19) Olorunniji, F. J., Buck, D. E., Colloms, S. D., McEwan, A. R., Smith, M. C., Stark, W. M., and Rosser, S. J. (2012) Gated rotation mechanism of site-specific recombination by varphiC31 integrase. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19661–19666.

(20) Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., Keasling, J. D., and Rosser, S. J. (2014) Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.* 42, e23.

(21) Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S., and Lu, T. K. (2016) Synthetic recombinase-based state machines in living cells. *Science (Washington, DC, U. S.)* 353, aad8559.

(22) Ghosh, P., Wasil, L. R., and Hatfull, G. F. (2006) Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol.* 4, e186.

(23) Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S., and Smith, M. C. (2011) A phage protein that binds phiC31 integrase to switch its directionality. *Mol. Microbiol.* 80, 1450–1463.

- (24) Mandali, S., Gupta, K., Dawson, A. R., Van Duyne, G. D., and Johnson, R. C. (2017) Control of Recombination Directionality by the Listeria Phage A118 Protein Gp44 and the Coiled-Coil Motif of Its Serine Integrase. *J. Bacteriol.* 199, e00019-17.
- (25) Pokhilko, A., Zhao, J., Ebenhoh, O., Smith, M. C., Stark, W. M., and Colloms, S. D. (2016) The mechanism of varphiC31 integrase directionality: experimental analysis and computational modelling. *Nucleic Acids Res.* 44, 7360–7372.
- (26) Fan, H. F., Hsieh, T. S., Ma, C. H., and Jayaram, M. (2016) Single-molecule analysis of varphiC31 integrase-mediated site-specific recombination by tethered particle motion. *Nucleic Acids Res.* 44, 10804–10823.
- (27) Olorunniji, F. J., McPherson, A. L., Rosser, S. J., Smith, M. C. M., Colloms, S. D., and Stark, W. M. (2017) Control of serine integrase recombination directionality by fusion with the directionality factor. *Nucleic Acids Res.* 45, 8635.
- (28) Bibb, L. A., Hancox, M. I., and Hatfull, G. F. (2005) Integration and excision by the large serine recombinase phiRv1 integrase. *Mol. Microbiol.* 55, 1896–1910.
- (29) Br  uner, A., Br  ndsted, L., and Hammer, K. (1999) Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901–1. *J. Bacteriol.* 181, 7291–7297.
- (30) Zhang, L., Zhu, B., Dai, R., Zhao, G., and Ding, X. (2013) Control of directionality in Streptomyces phage phiBT1 integrase-mediated site-specific recombination. *PLoS One* 8, e80434.
- (31) Fogg, P. C., Haley, J. A., Stark, W. M., and Smith, M. C. (2017) Genome Integration and Excision by a New Streptomyces Bacteriophage. *Appl. Environ. Microbiol.* 83, e02767-16.
- (32) Ramaswamy, K. S., Carrasco, C. D., Fatma, T., and Golden, J. W. (1997) Cell-type specificity of the Anabaena fdxN-element rearrangement requires xisH and xisI. *Mol. Microbiol.* 23, 1241–1249.
- (33) Abe, K., Kawano, Y., Iwamoto, K., Arai, K., Maruyama, Y., Eichenberger, P., and Sato, T. (2014) Developmentally-regulated excision of the SPbeta prophage reconstitutes a gene required for spore envelope maturation in Bacillus subtilis. *PLoS Genet.* 10, e1004636.
- (34) Boccard, F., Pernodet, J.-L., Friedmann, A., and Gu  rineau, M. (1988) Site-specific integration of plasmid pSAM2 in Streptomyces lividans and S. ambofaciens. *Mol. Gen. Genet.* 212, 432–439.
- (35) Kuhstoss, S., Richardson, M. A., and Rao, R. N. (1991) Plasmid cloning vectors that integrate site-specifically in Streptomyces spp. *Gene* 97, 143–146.
- (36) Andreas, S., Schwenk, F., K  ter-Luks, B., Faust, N., and K  hn, R. (2002) Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLP recombination in mammalian cells. *Nucleic Acids Res.* 30, 2299–2306.
- (37) Groth, A. C., Fish, M., Nusse, R., and Calos, M. P. (2004) Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* 166, 1775–1782.
- (38) Vrljicak, P., Tao, S., Varshney, G. K., Quach, H. N., Joshi, A., LaFave, M. C., Burgess, S. M., and Sampath, K. (2016) Genome-Wide Analysis of Transposon and Retroviral Insertions Reveals Preferential Integrations in Regions of DNA Flexibility. *G3: Genes, Genomes, Genet.* 6, 805.
- (39) Held, P. K., Olivares, E. C., Aguilar, C. P., Finegold, M., Calos, M. P., and Grompe, M. (2005) In vivo correction of murine hereditary tyrosinemia type I by phiC31 integrase-mediated gene delivery. *Mol. Ther.* 11, 399–408.
- (40) Chalberg, T. W., Genise, H. L., Vollrath, D., and Calos, M. P. (2005) phiC31 integrase confers genomic integration and long-term transgene expression in rat retina. *Invest. Ophthalmol. Visual Sci.* 46, 2140–2146.
- (41) Ishikawa, Y., Tanaka, N., Murakami, K., Uchiyama, T., Kumaki, S., Tsuchiya, S., Kugoh, H., Oshimura, M., Calos, M. P., and Sugamura, K. (2006) Phage phiC31 integrase-mediated genomic integration of the common cytokine receptor gamma chain in human T-cell lines. *J. Gene Med.* 8, 646–653.
- (42) Olivares, E. C., Hollis, R. P., Chalberg, T. W., Meuse, L., Kay, M. A., and Calos, M. P. (2002) Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat. Biotechnol.* 20, 1124–1128.
- (43) Ye, L., Chang, J. C., Lin, C., Qi, Z., Yu, J., and Kan, Y. W. (2010) Generation of induced pluripotent stem cells using site-specific integration with phage integrase. *Proc. Natl. Acad. Sci. U. S. A.* 107, 19467–19472.
- (44) Karow, M., Chavez, C. L., Farruggio, A. P., Geisinger, J. M., Keravala, A., Jung, W. E., Lan, F., Wu, J. C., Chen-Tsai, Y., and Calos, M. P. (2011) Site-specific recombinase strategy to create induced pluripotent stem cells efficiently with plasmid DNA. *Stem Cells* 29, 1696–1704.
- (45) Zhao, C., Farruggio, A. P., Bjornson, C. R., Chavez, C. L., Geisinger, J. M., Neal, T. L., Karow, M., and Calos, M. P. (2014) Recombinase-mediated reprogramming and dystrophin gene addition in mdx mouse induced pluripotent stem cells. *PLoS One* 9, e96279.
- (46) Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S., and Calos, M. P. (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell Biol.* 21, 3926–3934.
- (47) Scimmenti, C. R., Thyagarajan, B., and Calos, M. P. (2001) Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic Acids Res.* 29, 5044–5051.
- (48) Ortiz-Urda, S., Thyagarajan, B., Keene, D. R., Lin, Q., Fang, M., Calos, M. P., and Khavari, P. A. (2002) Stable nonviral genetic correction of inherited human skin disease. *Nat. Med.* 8, 1166–1170.
- (49) Chalberg, T. W., Portlock, J. L., Olivares, E. C., Thyagarajan, B., Kirby, P. J., Hillman, R. T., Hoelters, J., and Calos, M. P. (2006) Integration specificity of phage phiC31 integrase in the human genome. *J. Mol. Biol.* 357, 28–48.
- (50) Venken, K. J., He, Y., Hoskins, R. A., and Bellen, H. J. (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science (Washington, DC, U. S.)* 314, 1747–1751.
- (51) Bischof, J., Maeda, R. K., Hediger, M., Karch, F., and Basler, K. (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3312–3317.
- (52) Invitrogen by ThermoFisher Scientific (2017) *User Guide—Jump-In CHO-K1 Retargeting Kit: For Targeted Integration of Your Gene of Interest in Jump-In CHO-K1 Cells*, MAN0005450.
- (53) Invitrogen by ThermoFisher Scientific (2017) *User Guide—Jump-In GripTite HEK293 Retargeting Kit: For Targeted Integration of Your Gene of Interest in Jump-In GripTite HEK293 Cells*, MAN0005451.
- (54) Schlake, T., and Bode, J. (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* 33, 12746–12751.
- (55) Seibler, J., and Bode, J. (1997) Double-reciprocal crossover mediated by FLP-recombinase: a concept and an assay. *Biochemistry* 36, 1740–1747.
- (56) Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M., and Brown, W. (2013) Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.* 13, 87.
- (57) Xu, Z., and Brown, W. R. (2016) Comparison and optimization of ten phage encoded serine integrases for genome engineering in Saccharomyces cerevisiae. *BMC Biotechnol.* 16, 13.
- (58) Bateman, J. R., Lee, A. M., and Wu, C. T. (2006) Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. *Genetics* 173, 769–777.
- (59) Bateman, J. R., Johnson, J. E., and Locke, M. N. (2012) Comparing enhancer action in cis and in trans. *Genetics* 191, 1143–1155.
- (60) Huang, J., Zhou, W., Dong, W., Watson, A. M., and Hong, Y. (2009) From the Cover: Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. *Proc. Natl. Acad. Sci. U. S. A.* 106, 8284–8289.
- (61) Weng, R., Chen, Y. W., Bushati, N., Cliffe, A., and Cohen, S. M. (2009) Recombinase-mediated cassette exchange provides a versatile

platform for gene targeting: knockout of miR-31b. *Genetics* 183, 399–402.

(62) Venken, K. J., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., Carlson, J. W., Levis, R. W., Spradling, A. C., Hoskins, R. A., and Bellen, H. J. (2011) MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nat. Methods* 8, 737–743.

(63) Zhang, X., Koolhaas, W. H., and Schnorrer, F. (2014) A versatile two-step CRISPR-and RMCE-based strategy for efficient genome engineering in *Drosophila*. *G3: Genes, Genomes, Genet.* 4, 2409–2418.

(64) Zhu, F., Gamboa, M., Farruggio, A. P., Hippenmeyer, S., Tasic, B., Schüle, B., Chen-Tsai, Y., and Calos, M. P. (2014) DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res.* 42, e34.

(65) Farruggio, A. P., Chavez, C. L., Mikell, C. L., and Calos, M. P. (2012) Efficient reversal of phiC31 integrase recombination in mammalian cells. *Biotechnol. J.* 7, 1332–1336.

(66) Boeke, J. D., Church, G., Hessel, A., Kelley, N. J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V. W., Holt, L., Isaacs, F. J., Kuiken, T., Lajoie, M., Lessor, T., Lunshof, J., Maurano, M. T., Mitchell, L. A., Rine, J., Rosser, S., Sanjana, N. E., Silver, P. A., Valle, D., Wang, H., Way, J. C., and Yang, L. (2016) GENOME ENGINEERING. The Genome Project-Write. *Science (Washington, DC, U. S.)* 353, 126–127.

(67) Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S., and Boeke, J. D. (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* 477, 471–476.

(68) Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M., Cai, Y., Zeller, K., Agmon, N., Han, J. S., Hadjithomas, M., Tullman, J., Caravelli, K., Cirelli, K., Guo, Z., London, V., Yeluru, A., Murugan, S., Kandavelou, K., Agier, N., Fischer, G., Yang, K., Martin, J. A., Bilgel, M., Bohutski, P., Boulier, K. M., Capaldo, B. J., Chang, J., Charoen, K., Choi, W. J., Deng, P., DiCarlo, J. E., Doong, J., Dunn, J., Feinberg, J. I., Fernandez, C., Floria, C. E., Gladowski, D., Hadidi, P., Ishizuka, I., Jabbari, J., Lau, C. Y., Lee, P. A., Li, S., Lin, D., Linder, M. E., Ling, J., Liu, J., Liu, J., London, M., Ma, H., Mao, J., McDade, J. E., McMillan, A., Moore, A. M., Oh, W. C., Ouyang, Y., Patel, R., Paul, M., Paulsen, L. C., Qiu, J., Rhee, A., Rubashkin, M. G., Soh, I. Y., Sotuyo, N. E., Srinivas, V., Suarez, A., Wong, A., Wong, R., Xie, W. R., Xu, Y., Yu, A. T., Koszul, R., Bader, J. S., Boeke, J. D., and Chandrasegaran, S. (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science (Washington, DC, U. S.)* 344, 55–58.

(69) Shen, Y., Stracquadanio, G., Wang, Y., Yang, K., Mitchell, L. A., Xue, Y., Cai, Y., Chen, T., Dymond, J. S., Kang, K., Gong, J., Zeng, X., Zhang, Y., Li, Y., Feng, Q., Xu, X., Wang, J., Wang, J., Yang, H., Boeke, J. D., and Bader, J. S. (2016) SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* 26, 36–49.

(70) Knight, T. F. (2003) Idempotent Vector Design for Standard Assembly of BioBricks, In *MIT Synthetic Biology Working Group Technical Reports* 2003.

(71) Shao, Z., Luo, Y., and Zhao, H. (2011) Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* 7, 1056–1059.

(72) Smolke, C. D. (2009) Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol.* 27, 1099–1102.

(73) Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* 39, 5790–5799.

(74) Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39, e82.

(75) Shetty, R., Lizarazo, M., Rettberg, R., and Knight, T. F. (2011) Assembly of BioBrick standard biological parts using three antibiotic assembly. *Methods Enzymol.* 498, 311–326.

(76) Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., and Keasling, J. D. (2010) BglBricks: A flexible standard for biological part assembly. *J. Biol. Eng.* 4, 1.

(77) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3, e3647.

(78) Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juárez, P., Fernández-del-Carmen, A., Granell, A., and Orzaez, D. (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One* 6, e21622.

(79) Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs* 3, 38–43.

(80) Li, M. Z., and Elledge, S. J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* 4, 251–256.

(81) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.

(82) Shao, Z., Zhao, H., and Zhao, H. (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37, e16.

(83) Quan, J., and Tian, J. (2009) Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* 4, e6441.

(84) Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A., and French, C. E. (2014) PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res.* 42, e154.

(85) Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* 10, 1788–1795.

(86) Sasaki, Y., Sone, T., Yoshida, S., Yahata, K., Hotta, J., Chesnut, J. D., Honda, T., and Imamoto, F. (2004) Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system. *J. Biotechnol.* 107, 233–243.

(87) Cheo, D. L., Titus, S. A., Byrd, D. R., Hartley, J. L., Temple, G. F., and Brasch, M. A. (2004) Concerted assembly and cloning of multiple DNA segments using in vitro site-specific recombination: functional analysis of multi-segment expression clones. *Genome Res.* 14, 2111–2120.

(88) Zhang, L., Zhao, G., and Ding, X. (2011) Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. *Sci. Rep.* 1, 141.

(89) Invitrogen by Life Technologies (2003) *User Guide—Gateway Technology: A Universal Technology To Clone DNA Sequences for Functional Analysis and Expression in Multiple Systems*, MAN0000282.

(90) Wang, X., Tang, B., Ye, Y., Mao, Y., Lei, X., Zhao, G., and Ding, X. (2017) Bxb1 integrase serves as a highly efficient DNA recombinase in rapid metabolite pathway assembly. *Acta Biochim. Biophys. Sin.* 49, 44–50.

(91) Merrick, C. A., Wardrope, C., Paget, J. E., Colloms, S. D., and Rosser, S. J. (2016) Rapid Optimization of Engineered Metabolic Pathways with Serine Integrase Recombinational Assembly (SIRA). *Methods Enzymol.* 575, 285–317.

(92) Olorunniji, F. J., Merrick, C., Rosser, S. J., Smith, M. C. M., Stark, W. M., and Colloms, S. D. (2017) Multipart DNA Assembly Using Site-Specific Recombinases from the Large Serine Integrase Family. *Methods Mol. Biol. (N. Y., NY, U. S.)* 1642, 303–323.

(93) Siuti, P., Yazbek, J., and Lu, T. K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* 31, 448–452.

(94) Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013) Amplifying genetic logic gates. *Science (Washington, DC, U. S.)* 340, 599–603.

- (95) Manzoni, R., Urrios, A., Velazquez-Garcia, S., de Nadal, E., and Posas, F. (2016) Synthetic biology: insights into biological computation, *Integrative biology. quantitative biosciences from nano to macro* 8, 518–532.
- (96) Bonnet, J., Subsoontorn, P., and Endy, D. (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8884–8889.
- (97) Ham, T. S., Lee, S. K., Keasling, J. D., and Arkin, A. P. (2006) A tightly regulated inducible expression system utilizing the *flm* inversion recombination switch. *Biotechnol. Bioeng.* 1, 1–4.
- (98) Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., and Lichtman, J. W. (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 450, 56–62.
- (99) Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G., and Collins, J. J. (2009) Synthetic gene networks that count. *Science (Washington, DC, U. S.)* 324, 1199–1202.
- (100) Weinberg, B. H., Pham, N. T. H., Caraballo, L. D., Lozano, T., Engel, A., Bhatia, S., and Wong, W. W. (2017) Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* 35, 453–462.
- (101) Courbet, A., Endy, D., Renard, E., Molina, F., and Bonnet, J. (2015) Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci. Transl. Med.* 7, 289ra283.
- (102) Ham, T., Lee, S. K., Keasling, J. D., and Arkin, A. P. (2008) Design and Construction of a Double Inversion Recombination Switch for Heritable Sequential Genetic Memory. *PLoS One* 3, 1–4.
- (103) Hsiao, V., Hori, Y., Rothenmund, P. W., and Murray, R. M. (2016) A population-based temporal logic gate for timing and recording chemical events. *Mol. Syst. Biol.* 12, 869.
- (104) Bowyer, J. E., Hsiao, V., Wong, W. W., and Bates, D. G. (2017) Mechanistic modelling of a recombinase-based two-input temporal logic gate. *Engineering Biology* 1, 40–50.
- (105) Bowyer, J., Zhao, J., Subsoontorn, P., Wong, W., Rosser, S., and Bates, D. (2016) Mechanistic Modeling of a Rewritable Recombinase Addressable Data Module. *IEEE transactions on biomedical circuits and systems* 10, 1161–1170.
- (106) Subsoontorn, P., and Endy, D. (2012) Design and Analysis of Genetically Encoded Counters. *Procedia Computer Science* 11, 43–54.
- (107) Sheth, R. U., Yim, S. S., Wu, F. L., and Wang, H. H. (2017) Multiplex recording of cellular events over time on CRISPR biological tape. *Science (Washington, DC, U. S.)* 358, 1457.
- (108) Perli, S. D., Cui, C. H., and Lu, T. K. (2016) Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science (Washington, DC, U. S.)* 353, aag0511.
- (109) Kalhor, R., Mali, P., and Church, G. M. (2017) Rapidly evolving homing CRISPR barcodes. *Nat. Methods* 14, 195–200.
- (110) Farruggio, A. P., and Calos, M. P. (2014) Serine integrase chimeras with activity in *E. coli* and HeLa cells. *Biol. Open* 3, 895–903.
- (111) Akopian, A., He, J., Boocock, M. R., and Stark, W. M. (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8688–8691.
- (112) Gordley, R. M., Gersbach, C. A., and Barbas, C. F. (2009) Synthesis of programmable integrases. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5053–5058.
- (113) Klippel, A., Cloppenburg, K., and Kahmann, R. (1988) Isolation and characterization of unusual *gin* mutants. *EMBO J.* 7, 3983–3989.
- (114) Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H., and Liu, D. R. (2016) A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* 44, 9758–9770.
- (115) Ghosh, P., Pannunzio, N. R., and Hatfull, G. F. (2005) Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites. *J. Mol. Biol.* 349, 331–348.
- (116) Thorpe, H. M., Wilson, S. E., and Smith, M. C. (2000) Control of directionality in the site-specific recombination system of the *Streptomyces* phage phiC31. *Mol. Microbiol.* 38, 232–241.